

coverslip. When KB cells were grown on these composite coverslips and treated with directin, the orientation of the cells followed the longitudinal direction of the coverslip in a zigzag fashion. Experiments with different kinds of glass and other substrates are now in progress.

Summary.—We have found an agent in extracts derived from human urine which can be identified by its induction of directional growth of malignant cells in a defined *in vitro* system. The extract sensitizes malignant cells to submicroscopic, directional clues in their environment. Nine cell lines have been examined so far, four malignant and five normal ones. The normal cell lines studied did not respond to directin.

The authors wish to thank Drs. G. Moore, D. Stone, M. Steinberg, S. Inoue, A. Moscona, and R. D. Allen for discussions and suggestions. They thank Dr. D. Stone for permission to mention his preliminary results with directin.

* This research was supported by National Institutes of Health grant GM 10383.

† Present address: Battelle Memorial Institute, Columbus 1, Ohio.

¹ Weiss, P., *J. Exptl. Zool.*, **68**, 393–448 (1934).

² Weiss, P., *Anat. Record*, **124**, 381 (1956).

STUDIES ON POLYNUCLEOTIDES, XLIX. STIMULATION OF
THE BINDING OF AMINOACYL-sRNA'S TO RIBOSOMES BY
RIBOTRINUCLEOTIDES AND A SURVEY OF CODON
ASSIGNMENTS FOR 20 AMINO ACIDS†*

BY D. SÖLL, E. OHTSUKA, D. S. JONES, R. LOHRMANN, H. HAYATSU,
S. NISHIMURA, AND H. G. KHORANA

INSTITUTE FOR ENZYME RESEARCH OF THE UNIVERSITY OF WISCONSIN, MADISON

Communicated by Fritz Lipmann, June 23, 1965

Methods have been developed in this laboratory for the preparation of long ribopolynucleotides containing known repeating di- and trinucleotide sequences^{1, 2} and their use as messengers in the cell-free amino acid-incorporating system has permitted further studies of the amino acid code.^{3, 4} Thus, the ribopolynucleotide containing repeating ApApGp sequence directs the synthesis of the homopolypeptides of lysine, glutamic acid, and arginine;³ and the ribopolynucleotides, poly UC, poly UG, poly AC, and poly AG, containing in every case the two nucleotides in alternating sequence, direct the synthesis, respectively, of the copolypeptides of serine and leucine, of valine and cysteine, of threonine and histidine, and of arginine and glutamic acid.^{4, 5} These results have provided definitive answers to some of the basic characteristics of the amino acid code and have given information on the codon sequences for the amino acids involved in the polypeptide synthesis such that the assignment can be restricted to either two or three alternatives. For further work on the codon assignments for different amino acids, we have used the elegant technique of Nirenberg and Leder⁶ in which the effect of ribotrinnucleotides on the binding of specific aminoacyl-sRNA's to ribosomes is measured. From a combination of this technique and of the results of specific polypeptide synthesis,

the new codon sequence AAG⁷ for lysine and the codon sequence GAA⁷ for glutamic acid were deduced, and the sequence AGA was concluded for arginine despite the failure of the trinucleotide to stimulate the binding of arginyl-sRNA.³ In continuing this work, we have now carried out extensive testing of all of the possible ribotrinucleotides in the binding technique,⁶ and the present report describes the pertinent findings on the codon assignments for all of the 20 amino acids. In several recent publications, Nirenberg and co-workers have also presented results of their studies on the codon assignments by the binding technique.⁸⁻¹² The results of the binding studies herein reported, which are mostly in accord with the findings of Nirenberg and co-workers, are evaluated in relation to the codon assignments by independent methods. A brief review of most of the present work has been presented elsewhere.¹

Materials and Methods.—All of the ribotrinucleotides used in this study were prepared by chemical methods which will be described elsewhere.¹³ All these have 2'- and 3'-hydroxyl groups at one end and a 5'-hydroxyl group at the opposite end. Radioactive amino acids were purchased commercially. Their specific activities ($\mu\text{C}/\mu\text{mole}$) were as follows: C¹⁴-ala, 123; C¹⁴-arg, 246; C¹⁴-asn, 29.8; C¹⁴-asp, 164; H³-cys, 1000; C¹⁴-gln, 32.2; C¹⁴-glu, 205; C¹⁴-gly, 66; C¹⁴-his, 222; C¹⁴-ileu, 246; C¹⁴-leu, 198 and 222; C¹⁴-lys, 198; H³-met, 2000; C¹⁴-phe, 297; C¹⁴-pro, 165; C¹⁴-ser, 123; C¹⁴-thr, 164; H³-try, 3000; C¹⁴-tyr, 297; C¹⁴-val, 165.

sRNA was prepared from *Escherichia coli* B by the method of Zubay.¹⁴ Ribosomes and the crude mixture of aminoacyl-sRNA synthetases were prepared as described previously.³ Radioactive-labeled aminoacyl-sRNA's were prepared either from the total *E. coli* sRNA or from purified fractions by the method of Nishimura *et al.*³ with only one labeled amino acid, using 4-22 μmoles of amino acid/mg of total sRNA.

Fractionation of sRNA: For the preparation of labeled aminoacyl-sRNA's corresponding to the following amino acids, the unfractionated *E. coli* sRNA was used: ala, cys, ileu, leu, lys, phe, pro, ser, thr, tyr, and val. The preparation of the aminoacyl-sRNA's corresponding to the amino acids described below and additional preparations of aminoacyl-sRNA's corresponding to leu and ileu, lys and ser were made using fractionated sRNA. Fractionation of *E. coli* sRNA was performed by the technique of countercurrent distribution and the pattern of distribution of optical density as well as the location of acceptor activity peaks for aspartic acid, arginine, glutamic acid, and lysine have been described previously.³ The general order of location of the acceptor activities for the above amino acids and for the following amino acids in that distribution³ was similar to that of Goldstein, Bennett, and Craig.¹⁵ Asparagine acceptor activity was present in fractions 58-100, and material in fraction 78, the middle of the peak, was taken. Cysteine acceptor activity was present as a sharp peak in fractions 112-124 and fraction 118 was used. Glycine acceptor activity appeared as a single peak in fractions 52-88, and fraction 76 was used. Histidine acceptor activity was present in fractions 58-100, and fraction 88 was taken. Isoleucine acceptor activity gave three peaks, all being present in fractions 40-124. The three fractions (64, 88, and 118) selected are represented here as peaks I, II, and III, respectively. Leucine acceptor activity gave three peaks.¹⁶ Peak I, the major peak, was in fractions 100-124, and fraction 114 was taken. Peak III was present in fractions 174-190. These were combined for redistribution by the method of Apgar *et al.*¹⁷ In the present work, peak III is designated as that material which was insoluble in the redistribution solvent system. Peak IV is designated as that material which on redistribution of the soluble portion of peak III formed the major slower-moving peak. Methionine acceptor activity gave two peaks; the first, the major one, was present in fractions 64-94. Fraction 76 was taken.¹⁸ Serine acceptor activity gave two peaks, again in agreement with the results of Goldstein *et al.*¹⁵ Peak I was present in fractions 52-94, and fraction 76 was used. Peak II was in fractions 106-136, and fraction 118 was used. Tryptophane gave one peak and it was present in fractions 112-130, fraction 124 being used.

Aminoacyl-sRNA binding to ribosomes: The method of Nirenberg and Leder⁶ was used; the incubation mixtures (0.05 ml) contained buffer, 0.05 M KCl, 0.02 M magnesium acetate except where specified otherwise, 1.5-2.0 OD₂₆₀ units of ribosomes, and 0.2 OD₂₆₀ units of trinucleotides. The buffers used were 0.1 M tris-chloride, pH 7.5, or sodium cacodylate-HCl, pH 7.2, or tris-

Aminoacyl-sRNA (μ moles) ^a	Trinucleotide	Back-ground	+Trinucleotide	Other trinucleotides tested	Aminoacyl-sRNA (μ moles) ^a	Trinucleotide	Back-ground	+Trinucleotide	Other trinucleotides tested
C ¹⁴ -Gly ^b (24.4)	GGU GGC GGA GGG UGG CGG AGG	0.94 0.40 0.51 0.51 0.94 0.40 0.51	8.08 6.00 1.99 1.15 1.29 0.89 1.12	GUG, GCG, GAG, CUG, UGA, UAA, UAG, CUA, UGU	C ¹⁴ -Thr ^b (22.8)	ACA ACG ACU ACC CAC AAC	0.44 0.44 0.33 0.31 0.31 0.44	2.37 3.50 2.03 1.81 0.57 0.63	CCA, CAA, UGA, UAA, UAG, CUA, UGU, AGG, GGG
C ¹⁴ -His ^b (9.6)	CAC CAU CCA UCA	0.31 0.34 0.31 0.34	5.25 3.92 0.47 0.58	ACC, ACU, AUC, UAC, CUA, UGA, UAA, UAG, UGU, AGG, GGG	H ³ -Try ^b (12.7)	UGG GUG UGU AGG GGG	0.55 0.55 0.56 0.56 0.56	1.18 1.10 0.87 0.86 0.69	GGU, UGA, UAA, CUA, UAG
C ¹⁴ -Ileu ^c (25.2)	AUU AUC	0.24 0.25	1.17 1.82	UAU, UUA, UAG, CAU, ACU, UCA, CUA, UGA, UAA, UAG, UGU, AGG, GGG	C ¹⁴ -Tyr ^b (17.7)	UAU UAC CAU	0.33 0.33 0.33	2.79 3.16 0.65	AUU, UUA, AUC, ACU, UCA, CUA, UGA, UAA, UAG, UGU, AGG, GGG, CAC
					C ¹⁴ -Val ^b (41.4)	GUU GUG GUA GUC CGU UGU	0.32 0.32 0.32 0.32 0.32 0.35	1.97 7.76 7.74 2.55 0.66 1.24	UUG, GGU, UGG, UGA, UAA, UAG, CUA, AGG, GGG

^a The amounts (μ moles) of labeled aminoacyl-sRNA added in each 0.05 m incubation mixture are shown in parentheses under each aminoacyl-sRNA.

^b Buffer used, Tris-chloride, pH 7.5.

^c Buffer used, Na-acetate, pH 7.2.

^d Buffer used, Tris-acetate, pH 7.2.

^e This preparation of C¹⁴-asn-sRNA probably contained C¹⁴-asp-sRNA due to deamination of C¹⁴-asn prior to the charging reaction by the aminoacyl-sRNA synthetases [Schwartz, J. H., these PROCEEDINGS, 53, 1133 (1965)]. Because of low specific activity of C¹⁴-asn, the reaction mixture was scaled up to 0.2 ml. Each reaction mixture contained 20 μ moles (in base residues) of the trinucleotide, 0.02 M Mg⁺⁺, labeled aminoacyl-sRNA, and one of the buffers shown above. Incubations were at 25° for 20 min. Other details as in *Materials and Methods*. The binding results shown are for those trinucleotides which gave stimulation higher than 10% above background. "Other trinucleotides tested" refers to those trinucleotides which were also tested for the same amino acid but where the stimulation observed was less than 20% above background.

acetate, pH 7.2. The concentrations (μ moles/incubation mixture) of labeled aminoacyl-sRNA's used are shown for each aminoacyl-sRNA in Table 1. The incubations were at 25° for 20 min.

Results and Discussion.—Codon assignments from relatively specific stimulation of aminoacyl-sRNA binding: The data reported previously on amino acid incorporation using randomly linked ribopolynucleotides^{19, 20} and those obtained more recently on the synthesis of copolypeptides containing two amino acids in alternating sequence^{1, 4, 5} served as a guide to the selection of trinucleotides for testing the stimulation of the different aminoacyl-sRNA's to ribosomes. The results now obtained on the stimulation of the binding to ribosomes of the aminoacyl-sRNA's corresponding to the 20 amino acids are listed in Table 1. Shown in this table are groups of trinucleotides, which gave detectable (20% or more above the background) stimulation, and of those trinucleotides which were tested but gave no stimulation of the binding of a particular aminoacyl-sRNA. As is seen from the table, different trinucleotides varied greatly in their template activity. Many trinucleotides caused substantial and relatively specific stimulation of the binding of only one aminoacyl-sRNA to ribosomes and in these cases codon assignments can be made with considerable certainty. Many of the assignments included in Table 2 were thus made from the results of binding studies alone. Table 1 also shows further that, often, an additional number of trinucleotides caused relatively weak stimulation of the binding of an aminoacyl-sRNA. Thus, for example, in addition to the four trinucleotides, GCC, GCG, GCA, and GCU, which are concluded to be the codons for alanine (Table 2), the trinucleotide, AGC, also stimulated the binding of ala-sRNA to ribosomes. Similar examples of this kind are abundant for many of the other aminoacyl-sRNA's. These are omitted from consideration as codons, because, most often, each one of these trinucleotides stimulated more significantly the binding of a different aminoacyl-sRNA.

Ambiguity and lack of specificity in the stimulation of aminoacyl-sRNA binding by trinucleotides: In addition to the type of situation discussed above, there were cases where one trinucleotide stimulated significantly the binding of more than one aminoacyl-sRNA and the results were difficult to reconcile with the evidence obtained by independent approaches. In some cases the ambiguity observed in binding experiments at 25° in 0.02 M Mg^{++} ion concentration could be overcome by using 37° and 0.01 M Mg^{++} ion concentration (cf. ref. 3). For example, under the standard conditions of Table 1, GCG, an alanine codon, and CCG, a proline codon, both stimulated the binding of arg-sRNA and their template activity was higher than that of CGG for this aminoacyl-sRNA. However, at 37° and 0.01 M Mg^{++} ion concentration, the stimulation of arg-sRNA binding by GCG and by CCG was depressed relative to CGG and the latter codon, along with CGA and CGC, are concluded for arginine (Tables 1 and 2).

GGA stimulated the binding of gly-sRNA to ribosomes but it also stimulated the binding of glu-sRNA, the effect in the latter case being comparable to that given by GAG, a codon for glutamic acid.^{1, 3, 5} While the assignment of GGA as a codon for glycine is consistent with the results of Yanofsky,²¹ the ambiguity shown by this trinucleotide for glu-sRNA binding could not be diminished by the use of higher temperature and lower Mg^{++} ion concentration.

Stimulation of the binding of C¹⁴-gly-sRNA and C¹⁴-phe-sRNA was given by the trinucleotide AGG. No amino acid has in fact been assigned to this trinucleotide in

TABLE 2
CODON ASSIGNMENTS FROM POLYPEPTIDE SYNTHESIS AND/OR STIMULATION OF
AMINOACYL-SRNA BINDING TO RIBOSOMES*

1st	U	C	2nd	A	G	3rd
U	<u>Phe</u>	<u>Ser</u>		Tyr	<u>Cys</u>	U
	<u>Phe</u>	<u>Ser</u>		Tyr	<u>Cys</u>	C
		Ser				A
C	Leu	Ser			Try	G
		Pro		His	Arg	U
	<u>Leu</u>	<u>Pro</u>		<u>His</u>	<u>Arg</u>	C
A		<u>Pro</u>		<u>Gln</u>	Arg	A
	Leu	Pro		Gln	Arg	G
	Ileu	Thr		Asn	Ser	U
G	Ileu	Thr		Asn	Ser	C
		<u>Thr</u>		<u>Lys</u>	<u>Arg</u>	A
	Met	<u>Thr</u>		<u>Lys</u>		G
G	Val	Ala		<u>Asp</u>	Gly	U
	Val	Ala		Asp	Gly	C
	Val	Ala		<u>Glu</u>	Gly	A
	<u>Val</u>	Ala		<u>Glu</u>	Gly	G

* This method of presentation of the codon sequences originated with Dr. F. H. C. Crick (private communication).

The assignments not underlined are on the basis of binding experiments only. The assignments singly underlined are on the basis of copolypeptide and/or homopolypeptide syntheses and gave essentially no binding. The assignments doubly underlined are derived from both polypeptide synthesis and binding experiments.

Table 2, since the binding of gly-sRNA was abolished at higher temperature and the stimulation of phe-sRNA was most probably an artifact as judged by the incorporation data using randomly linked polynucleotides.^{19, 20}

Lack of stimulation of the binding of aminoacyl-sRNA's by certain trinucleotides: At least five cases were encountered in which the trinucleotides, which are believed to be authentic codons from other lines of work, failed to give any detectable stimulation of the binding of the appropriate aminoacyl-sRNA's under a variety of conditions that were tested. In addition, there were cases where the stimulation of binding observed was not sufficiently strong to justify codon assignment on the basis of this technique. Thus, in agreement with the results of Bernfield and Nirenberg,¹⁰ CUC and CUU failed to give any stimulation of the binding of leu-sRNA. The assignment of CUC codon for leucine is, nevertheless, assured from the synthesis of seryl-leucine copolypeptide under the direction of poly UC containing U and C in alternating sequences,⁴ and there is now a strong case for the assignment of CUU codon also for leucine from the work on acridine mutants of phage T4 lysozyme.^{22, 23} The same work²² also strongly suggests the assignment of UUA as another codon for leucine but this trinucleotide also failed to give convincing stimulation of the binding of leu-sRNA, a weak stimulation being observed when leu-sRNA from peak IV was tested (Table 1).

The assignment of AGA as a codon for arginine was concluded from the results of homo- and copolypeptide syntheses,^{1, 3, 5} and the assignment is in agreement with the data on the amino acid replacements obtained by Yanofsky and co-workers.²¹ No stimulation of the binding of arg-sRNA to ribosomes was given by AGA.

AGU can now be concluded to be a codon for serine²² and AGC would also appear to be a codon for the same amino acid.²⁴ These two trinucleotides, however, gave relatively weak stimulation of the binding of ser-sRNA. (The stimulation by UCU,

UCC, UCA, UCG, which are all concluded to be codons for serine, was more marked; see Table 1.)

AUA, which has been predicted as a codon for isoleucine¹⁰ or methionine,¹² gave relatively weak stimulation of the binding of aminoacyl-sRNA's corresponding to both of the amino acids and the assignment on the basis of these results cannot be conclusive. CUA was tested for the stimulation of the binding of all of the aminoacyl-sRNA's and only leu-sRNA gave weak (50% above background) response. No assignment has been made for this trinucleotide. Similarly, no assignment can be made for UGA, which gave slight stimulation of the binding of several aminoacyl-sRNA's under the standard conditions of Table 1 but did not give any stimulation at 37° and 0.01 M Mg⁺⁺ ion concentration. Finally, UAA and UAG, which have been concluded to be chain-terminating triplets^{25, 26} were tested for the binding of all of the 20 aminoacyl-sRNA's and found to give no detectable stimulation. While these negative results are consistent with the conclusions from the genetical experiments, they cannot constitute strong support for them since as discussed above authentic codon triplets can fail to stimulate the binding of appropriate aminoacyl-sRNA's.

General Comments.—Of the assignments shown in Table 2, maximum confidence is placed on those which are supported both by polypeptide synthesis and by the binding experiments (these assignments are doubly underlined). The assignments singly underlined in Table 2 are on the basis of results obtained in cell-free polypeptide synthesis or other independent methods. These cases have already been discussed above. The remainder of the assignments (not underlined) are on the basis of binding experiments alone and are mostly in accord with the results of Nirenberg and co-workers.¹² Remarkably useful and elegant as this technique is for codon assignments, it is important to emphasize at this stage that it is highly desirable to obtain independent evidence on these assignments. As noted above, there are examples where "authentic" triplets bind only poorly or not at all and, on the other hand, there are examples of unspecific or ambiguous binding. It is entirely possible that the mechanism of aminoacyl-sRNA binding as stimulated by trinucleotides is less stringent or only partially analogous to that of binding in the presence of messenger-like long ribopolynucleotides.

Extensive degeneracy in the amino acid code may now be regarded as firmly established and there appear to emerge a few general patterns. In one type of degeneracy, the third letter may be either of the two purines. Thus, GAA and GAG both code for glutamic acid and AAG and AAA both code for lysine. Similarly, the third letter may be either of the two pyrimidines. Thus, UUU and UUC both code for phenylalanine and there are many examples of this type of degeneracy in Table 2. In the third type of degeneracy, the third letter may be any one of the four bases. Several examples of this type are seen in Table 2. However, further work is necessary to evaluate the extent of this type of degeneracy. Finally, there are cases of degeneracy where the first letter may be exchanged. Thus, AGA and CGA both code for arginine and CUG and UUG both code for leucine.

Summary.—All of the possible trinucleotides have been tested extensively for their ability to stimulate the binding of labeled aminoacyl-sRNA's to ribosomes. While a large number of trinucleotides stimulated the binding of only one aminoacyl-sRNA to ribosomes significantly, many cases of ambiguous as well as weak

stimulation of the binding of aminoacyl-sRNA's were encountered. In addition, five trinucleotides, concluded from other methods to be authentic codons, failed to stimulate the binding of the appropriate aminoacyl-sRNA. By a combination of the results of the binding technique and of the results obtained by other methods, a total of 56 codon assignments have been made.

* Paper XLVIII is by S. Nishimura, D. S. Jones, and H. G. Khorana [*J. Mol. Biol.*, **13**, 302 (1965)].

† This work has been supported by grants from the National Cancer Institute of the National Institutes of Health, U.S. Public Health Service (grant no. CA-05178), National Science Foundation (grant no. GB-976), The Life Insurance Medical Research Fund (grant no. G-62-54), and the Wisconsin Alumni Research Fund.

¹ Khorana, H. G., *Federation Proc.*, in press.

² Khorana, H. G., T. M. Jacob, M. W. Moon, S. A. Narang, and E. Ohtsuka, *J. Am. Chem. Soc.*, **87**, 2954 (1965).

³ Nishimura, S., D. S. Jones, E. Ohtsuka, H. Hayatsu, T. M. Jacob, and H. G. Khorana, *J. Mol. Biol.*, **13**, 283 (1965).

⁴ Nishimura, S., D. S. Jones, and H. G. Khorana, *J. Mol. Biol.*, **13**, 302 (1965).

⁵ Jones, D. S., S. Nishimura, and H. G. Khorana, unpublished work.

⁶ Nirenberg, M. W., and P. Leder, *Science*, **145**, 1399 (1964).

⁷ In the specification of the trinucleotides, the system of abbreviations in current use is further simplified, for convenience, by omitting the letter *p* between the nucleoside initials; thus, for example, ApApG is simply shown as AAG, GpApA is shown as GAA, and so on.

⁸ Leder, P., and M. W. Nirenberg, these PROCEEDINGS, **52**, 420 (1964).

⁹ *Ibid.*, p. 1521.

¹⁰ Bernfield, M. R., and M. W. Nirenberg, *Science*, **147**, 479 (1965).

¹¹ Trupin, J. S., F. M. Rottman, R. L. C. Brimacombe, P. Leder, M. R. Bernfield, and M. W. Nirenberg, these PROCEEDINGS, **53**, 807 (1965).

¹² Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal, these PROCEEDINGS, **53**, 1161 (1965).

¹³ Lohrmann, R., D. Söll, E. Ohtsuka, H. Hayatsu, and H. G. Khorana, unpublished work.

¹⁴ Zubay, G., *J. Mol. Biol.*, **4**, 347 (1962).

¹⁵ Goldstein, J., T. P. Bennett, and L. C. Craig, these PROCEEDINGS, **51**, 119 (1964).

¹⁶ Weisblum, B., S. Benzer, and R. W. Holley, these PROCEEDINGS, **48**, 1449 (1962).

¹⁷ Apgar, J., R. W. Holley, and S. H. Merrill, *J. Biol. Chem.*, **237**, 796 (1962).

¹⁸ The results obtained from testing the material of the second peak with the trinucleotides AUA and AUG were similar to those using material from the major peak and are not shown in Table 1.

¹⁹ Ochoa, S., *Federation Proc.*, **22**, 62 (1963).

²⁰ Nirenberg, M. W., J. H. Matthaei, O. W. Jones, R. G. Martin, and S. H. Barondes, *Federation Proc.*, **22**, 55 (1963).

²¹ Yanofsky, C., *Biochem. Biophys. Res. Commun.*, **18**, 898 (1965).

²² Okada, Y., and G. Streisinger, unpublished work.

²³ It may also be noted that while CUU fails to give any stimulation, CUUUU was recently reported to give good stimulation of the binding of leucyl-sRNA to ribosomes. Thach, R., T. A. Sundararajan, and P. M. Doty, *Federation Proc.*, **24**, 409 (1965).

²⁴ Wittmann, H. G., in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 28 (1963), p. 589.

²⁵ Brenner, S., A. O. W. Stretton, and S. Kaplan, *Nature*, **206**, 994 (1965).

²⁶ Weigert, M. G., and A. Garen, *Nature*, **206**, 992 (1965).